

THE ISL MEETING CORPUS: CATEGORICAL FEATURES OF COMMUNICATIVE GROUP INTERACTIONS

Susanne Burger, Zachary A Sloane

Interactive Systems Laboratories, Carnegie Mellon University, Pittsburgh, PA, USA
sburger@cs.cmu.edu, zzs+@pitt.edu

ABSTRACT

Current research in human discourse structure has to this point focused heavily on two-party dialogues and established several dialogue annotation schemata. This study examines whether the HCRC Dialogue Structure Coding Scheme (DSCS) [1] provides an appropriate tool to express structures of communicative group interactions in multiparty conversations. We examine a test set of multiparty meetings—recorded conversation sessions involving more than two speakers—using a derivative of DSCS, which was originally designed for two-speaker dialogue situations.

DSCS provides a three-level system, which annotates communicative acts as *discourse moves* according to their discourse goals. It also captures the higher-level structure of dialogues as *games* and then combines games into *transactions*.

This preliminary study examines five different meeting types defined in the ISL Meeting Corpus. The speech turns from each sample meeting are annotated as discourse moves, which in sequences comprise discourse games, using an adapted form of the DSCS.

The initial results suggest that it is possible to adapt the DSCS scheme to multiparty conversations of different type. The data show that several distinct structural features exist in subsections of meetings and across different meeting types. Further, the results indicate that DSCS is useful for analyzing speaker involvement and participant hierarchies within a conversation. DSCS provides a means for discovering reoccurring patterns in meetings and isolating higher-level structures containing information units for discourse analysis and automatic recognition systems training.

1. INTRODUCTION

Meetings differ in organizational structure, speaking style, duration and result, depending on the participants and the task. Meetings are not dialogues; they are ‘polylogues’ or multiparty conversations. This makes communication

more complex because speakers may address more than one person at a time, and more than one person may react to a single communicative act.

However, as with dialogues, multiparty conversations consist of a series of communicative events, which can be generally defined as any act by which one person gives or receives information about that person's needs, desires, perceptions, knowledge, or affective states.

Although these communication acts will be more complexly interwoven among more speakers, this parallel between dialogues and meetings is promising for an effort to find a reasonable ontology and description scheme for meetings.

Our study applies the HCRC Dialogue Structure Coding Scheme (DSCS) (Carletta et al, 1996) to a selection of meetings from the ISL Meeting Corpus to see if it is useful to extend a scheme developed for dialogue annotation to multiparty conversation for the purpose of analyzing meeting structure and participants behavior and relationship.

2. THE ISL MEETING CORPUS

The ISL Meeting Corpus consists of audio recordings of 112 meetings, totaling about 107 hours of conversation. Meetings typically have between three and 12 participants, and average about six. 50% of these meetings can be shared with other research groups. For these meetings, all participants signed permission sheets for release. 18 of the shareable meetings will be available at the Linguistic Data Consortium; the others are being transcribed.

2.1. Technical setup

During meeting recordings, each speaker wore an individual lapel microphone and was recorded via multi-channel mix board and multi-channel sound card. This setup was devised to obtain a consumer- or application-style sound quality.

Many of the recordings also used various experimental setups in parallel to the lapel microphone setup, depending on the number of available recording channels. These

additional setups included wireless lapel microphones, headset microphones, table microphones, and a ceiling-mounted microphone array. For 46 meetings video was also recorded using up to three handheld cameras, a panoramic camera, or a speaker-tracking system.

2.2. Recording environment

All meetings were recorded in an open-plan office and lab environment with the typical background noises and artificial light. The meeting area is separated from the larger office area by three cubicle walls. Participants sat around an oval table, and sometimes made use of a smart board, white board, wall projectors or a TV.

2.3. Participants

Participants were project partners, groups from other labs, students and co-workers. Most of the participants spoke English natively; others were non-native speakers of English. English is the only language used in the meetings.

2.4. Scenarios

The recorded meetings were either *natural* meetings where participants needed to meet in the real world, or *artificial* meetings, which were designed explicitly for the purposes of data collection.

The *natural* meetings were work-related; the participants had either scheduled a meeting in the meeting space of the recording lab or had been invited there. The meeting agenda was always real, unrelated to the ISL recording, and known beforehand.

The *artificial* meetings provided topics to the participants. The topic could be a controversial discussion subject (controversial subjects were used to elicit the most active discussion), or it could be an open-ended instruction to ‘just chat’ about whatever came to mind. Participants were also given games to play: board games, card games, and role-playing all appear in the ISL corpus. Typically, participants needed to solve problems, answer questions or role-play by acting out characters in a made-up situation.

The meetings of the ISL Meeting Corpus are organized according to the meeting scenario and sorted into different meeting types, where *project* and *work* meetings use a natural scenario and *chatting*, *discussion* and *game playing* use artificially created scenarios.

Table 1 shows a broad description of the meeting types. See also [2] for more detail.

2.5. Transcription

The meetings were transcribed at the orthographic word level. In addition to words the transcriptions label spontaneous phenomena and disfluencies.

project meeting	Participants:	Project teams working on parts of a larger project
	Hierarchy:	Team leaders, team members
	Speaking Style:	Slow dynamic (few turns per minute and words per minute), many very long turns and very short turns
	Vocabulary:	Domain-dependent (one topic)
Work meeting	Participants:	Individual team members
	Hierarchy:	Supervisor and team members
	Speaking Style:	Slow dynamic, many very long turns and very short turns
	Vocabulary:	Domain-dependent (several topics)
Discussion	Participants:	Individuals
	Hierarchy:	Pro and con positions, speaker alliances, active speakers, eager speakers
	Speaking Style:	Fast dynamic (many turns per minute and words per minute), many very short turns
	Vocabulary:	Topic-centered, open domain
Chatting	Participants:	Individuals
	Hierarchy:	Balanced
	Speaking Style:	Fast dynamic, laughing, few of the very short and very long turns
	Vocabulary:	Open domain
Game playing	Participants:	Individuals and loose alliances
	Hierarchy:	Game-dependent
	Speaking Style:	Laughing, few of the very short and long turns
	Vocabulary:	Topic (game)-dependent

Table 1: Meeting types in the ISL Meeting Corpus

3. EXPERIMENT

The experiment attempted to elucidate discourse units in five meetings, one from each of the five types described above. Using the transcriptions of these meetings, which are divided into speech turns, the function of each discourse contribution by each speaker was annotated according to a simple labeling schema. These annotations were then segmented into higher-level sections reflecting patterns of initiation and response.

Afterwards, the annotated data was analyzed, with the focus on several preliminary questions:

- Are dialogue discourse-function labels sufficient to display the discourse dynamics of multiparty conversations? Is more refinement required?
- How is the larger number of participants reflected in the discourse patterns? That is, what do the patterns say about who was involved most in the conversation, and how many participants contributed to a discourse unit?

- Are there differences across meeting types with respect to the types of discourse labels used?

3.1. Test data

Five of the ISL meetings served as test data for the preliminary study. Each was selected from one of the meeting types—project, work, discussion, chatting and role-playing. The meetings are short in duration, to make labeling feasible, and have a large number of participants, to get a good variety of speakers and speaker interactions.

- **Project:** Several project teams of an integrated project met and planned an upcoming demonstration of their applications for customers. (702 turns, 12 speakers, 52 minutes)
- **Work:** In absence of the main supervisor, a work team met to discuss work status and distribute the workload for the next weeks. (697 turns, 6 speakers, 36 minutes)
- **Role-playing:** Male members of the lab met and pretended to be executives at a company; the task was to decide which representative car to purchase for the company vehicle. (1056 turns, 5 speakers, 27 minutes)
- **Discussion:** Members of the lab were told to discuss the topic ‘gun ownership.’ (193 turns, 5 speakers, 8 minutes)
- **Chatting:** Members of the lab were given a time limit and asked to kill time by chatting with each other about whatever came to mind. (383 turns, 4 speakers, 14 minutes)

3.2. Dialogue structure annotation schema

In order to label the discourse structure of the test meetings, the Dialogue Structure Coding Scheme (DSCS) from the Human Communication Research Centre (HCRC) ([1], [3]) was chosen for the following reasons: it is task independent and, therefore, extensible and adaptable. It specifies a manageable number of labels, labels can be assigned using an understandable decision tree, the system is learnable in a reasonable amount of time, and finally, we already used it in previous work [4].

DSCS attempts to both classify single utterances according to their discourse purpose into ‘moves’ and to capture the higher-level structure of a discourse in terms of the so-called ‘game’ structures. Moves are different kinds of initiations or responses classified according to their conversational function. A game is a set of turns starting with an initiation followed by responding moves until the purpose of the game is either fulfilled or abandoned.

3.3. Annotation procedure

<i>Move</i>	<i>Explanation</i>
1. Initiation Moves	Often occur at the beginning of a game where they set up a new series of moves
Ready	Signals initiation of a new game (e.g. “well”, “okay”)
Instruct	Direct instruction, command
Statement	Information of any type that is not elicited by another speaker
Align	Request for confirmation of successful communication (e.g. “Okay? Right?”)
Check	Request for confirmation of some mutually understood information
Query-YN	Yes/no question
Query-W	Open or WH-question
2. Response Moves	Used after initiation to fulfill the discourse expectations
Acknowledge	Confirmation, attention, feedback
Clarify	Additional information, explaining a response
Reply-Y	Affirmative answer
Reply-N	Negative answer
Reply-W	Complex answer
Added Moves	Moves that were added to the DSCS
Proposal	Direct proposal of action or offer
Correct	Direct correction or restatement of information
Contra	Statement that contradicts an initial statement
Polite	Greetings, please and thank you, etc
Out	Speech that is not part of the meeting conversation (e.g. talking on a cell phone)
Cough	Coughing
Laugh	Laughing
Throat	Clearing the throat
Noise	Any other human noise
Unknown	Turn is incomplete or otherwise impossible to label

Table2: Adapted Move Schema

The transcriptions of the ISL meetings are organized into successive speaker contributions, so-called turns. The order of turns is determined by the time where a speaker starts to talk. In cases of interfering turns, the turn, which started first is transcribed first and all other turns follow, even if they were produced during the first turn.

A turn may contain multiple discourse moves. The test meetings were made up of 3031 total speech turns; 464 of these had to be split to contain single moves. A split turn contained an average of 2.18 splits, with a maximum of four splits per turn.

The labeling procedure involved two human annotators. As a test set, both annotators in parallel annotated about

100 turns of each meeting. Differences in annotations were discussed to obtain consistency. It was decided to remove some label choices from the label set that were confusing or never used, and to add some test moves (*Proposal*, *Correct*, *Contra*, *Polite*, *Out*) to see if they would be useful and in which positions they would occur (initiation or response). Turns could contain also non-verbal events such as coughing or laughing. These were kept and labeled according to their content. Some turns were incomplete and could not be labeled at all. These were tagged as *unknown*. All together, the adapted annotation schema is a set of 22 discourse moves, including the ‘noise’ labels and the ‘unknown’ label, twelve moves from the DSCS, and five experimental moves (see table 2). Eventually, all five meetings were completely annotated according to the adapted schema.

After additional checking and correction passes, sets of moves were grouped into games by deciding whether a particular move initiated a new discourse unit or responded to a previous initiation. The first and last moves of each game were marked on an additional tier.

4. RESULTS

In total, the meetings contained 3617 discourse moves comprising 726 games. The games averaged 5.4 moves each, with the largest game containing 22 moves and the smallest games containing a single move.

4.1. Speaker involvement

4.1.1. How many speakers are involved in a game?

The number of speakers that may be involved in a game is limited only by the number of meeting participants, but in most of the games only a portion of the speakers are involved.

Games in which only one speaker was involved, or where none of the other speakers responded, were considered broken games. These games usually consist of one or two moves, and may include a speaker’s response to his own initiation. Only the role-play meeting had a significant frequency of these games, 5.8%.

Figure 1 shows the following results:

56% of the games in the *project* meeting involved only two speakers. 31% involved three speakers, 9% had four speakers, and 2% had five speakers.

41% of the games in the *work* meeting involved three speakers; 36%, two speakers; 17%, four speakers; and 3%, five speakers.

The role-playing meeting had more balanced speaker participation in its games: 35% of the games featured four speakers; 25%, three speakers; and 17%, two speakers. All five speakers are present in 16% of the games.

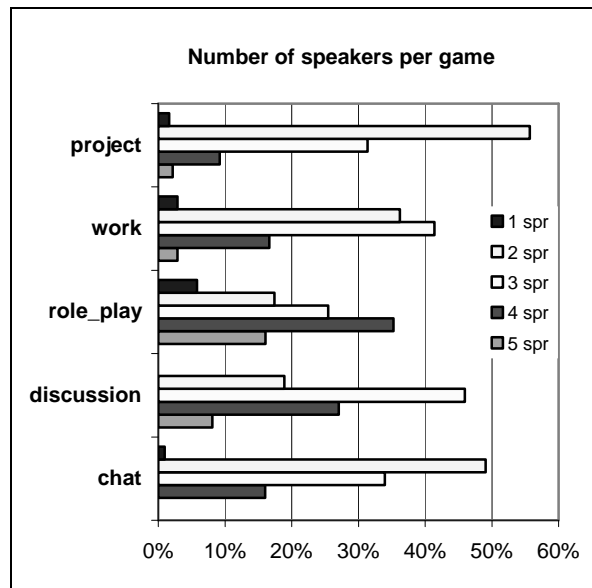


Figure 1: Percentage of speaker participation per game and meeting type

The discussion has three participants in 45% of the games, four speakers in 27% of the games, two speakers in 19% and all five participants together in 8%.

Finally, in the chat meeting, 49% of the games include only two speakers; 34% had three speakers and 16% had four speakers.

4.1.2. How often are individual speakers involved in games?

Figure 2 shows the results among the different types of games:

In the project meeting, two speakers clearly dominated: one spoke in 76% of the games, the other in 73%. Three other speakers appear in about a quarter of the games each (27%, 26%, and 19%, respectively); a sixth speaker participated in 135 of the games. The remaining six speakers appear in under 10% of the games.

The work meeting also had two dominant, equal speakers who appeared in just over 75% of the games, as well as a third speaker who was less dominant but still very involved (67%). The other three participants spoke in less than a third of the games (30%, 19% and 12%, respectively).

The role-play meeting had a balanced participation of all speakers (appearing in 71%, 70%, 70%, 67%, and 60% of games). There were no clear dominant speakers.

The discussion meeting had three speakers that appeared in more than half the turns, ranging from 92% for the primary dominant speaker, to 84% for the secondary dominant speaker, to 68% for the tertiary dominant speaker. The remaining two speakers still participated in 43% and 38% of the games.

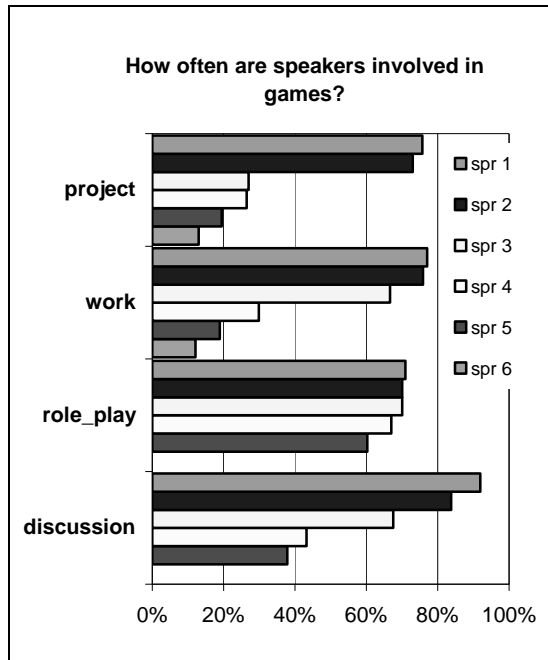


Figure 2: Percentage of game involvement of individual speakers of a meeting

Lastly, the chat meeting again had two dominant speakers (85% and 80%) and two speakers who participated in at least the half of the games (57% and 43%).

These results show differences in speaker participation among the meeting types. Two meetings, the project and work meeting, demonstrate a hierarchical structure including main speakers, speakers who speak frequently only at certain times, and speakers who contribute relatively little to the meetings.

In case of the project meeting, the dominant speakers monopolize the discussion, and the other participants contribute when called upon. More than half of the games in the project meeting have the same two speakers involved. These two speakers are the team leaders; the other participants are support team members.

We expected the work meeting to be similar: one main speaker—the supervisor—and other speakers contributing primary to specific topics within the meeting. In the meeting chosen for this experiment, however, the actual team supervisor was not present. Instead, the participation pattern reflects two ‘deputies’ competing for attention, along with another senior worker concurring with them. This seems to explain the fact that almost half the games involve the same three participants.

Chatting and discussion show significant involvement of all the participants, but at different frequencies. Both have at least two main speakers. The difference between the two meetings is the number of speakers usually involved in a game: in the discussion meeting, the most common games involve three or four speakers (though not

necessarily the same three or four in each game), while in the chatting meeting, more than half the games involve only two speakers, and still a third of the games involve three speakers.

The role game differs from the rest of the meetings in that both the levels of contribution per speaker as well as the number of speakers per game are nearly balanced. Each speaker contributed to a high percentage of games, and a high percentage of games involved four or even five speakers.

4.2. Move Patterns

4.2.1. Begin and End moves

Figure 3 shows how often certain move types started a game and how often certain moves ended a game: 40% of the beginning moves were statements, 25% were queries, 10% were ready moves, 7% were proposals and 6% were check moves. 52% of the end moves were acknowledgements, 13% were replies, 8% were clarifications and 6% were laughter. This distribution is in line with the DSCS specification of game structure. However, 9% of end moves were also statements, and 3% of all games started with a clarify move, which is technically a response move.

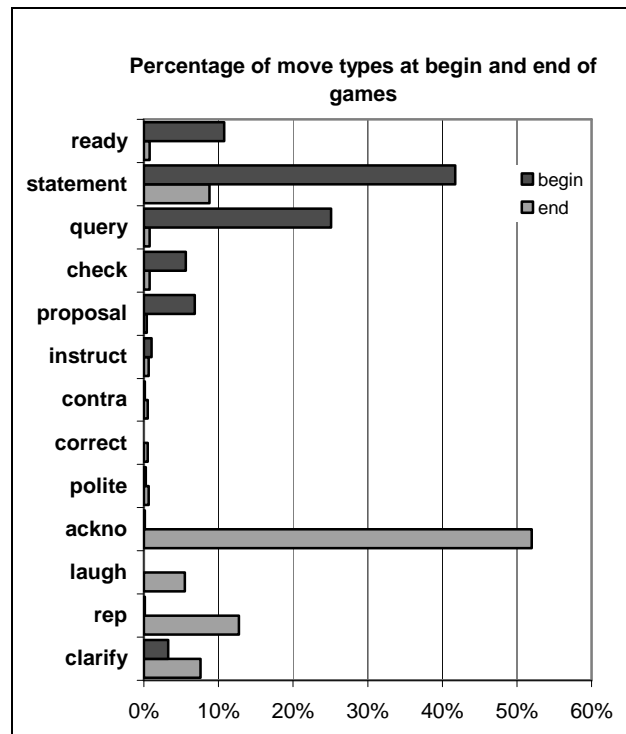


Figure 3: Percentage of moves beginning a game and moves ending a game

The statements that were found at the end of games contained mostly a summarization of the moves before or concluded the game.

The cases where clarify moves started a game are uncertain; in these cases 'statement' might have been used instead.

Since several moves were added to the DSCS scheme, it is interesting to note how they are distributed. These moves, *polite*, *correct*, and *contra*, occurred very seldom at the beginning or end of a game. Of those that did appear as game boundaries, end moves were slightly more prevalent.

The moves between the beginning and end of a game were called 'mid' moves. Overall, mid moves break down as follows: 39% acknowledgements, 17% replies, 10% statements, 9% clarification, 6% laughter, 5% queries, and 4% checks. Other moves can be found too, but at a frequency of less than 1%.

4.2.2. Move distribution in meeting types

The different meeting types differ with respect to both distribution and frequency of move label types.

Figure 4 shows patterns of the frequency of certain move types in percentage (move types with a frequency under 1% are omitted, the most frequent move *acknowledgement* is not displayed).

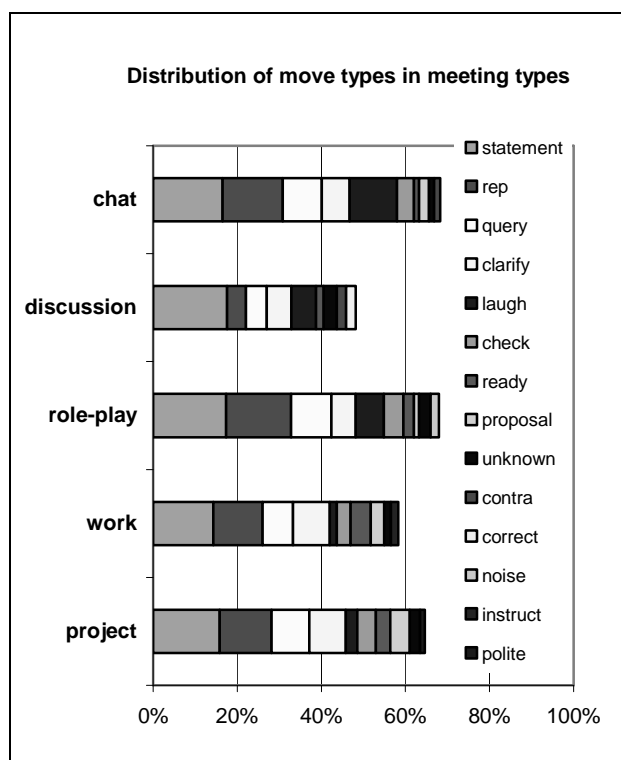


Figure 4: Move distribution in meeting types in percentage (move types with a frequency under 1% are omitted, the most frequent move *acknowledgement* is not displayed)

For all meeting types, acknowledgments were the most frequently annotated moves, at their highest rates in the work and discussion meetings (41% and 50%, respectively). Statements appear with consistent frequency (between 14% and 16%) in all meeting types. The meetings had similar frequencies of queries (around 9% of the moves) and replies (around 13% of the moves), with the exception of the discussion meeting, which had significantly fewer queries (5%) and replies (4.5%). The move type 'clarify' occurred in all meetings at an average frequency of 7%, 'ready' at an average of 3% and 'unknown' at an average of 2%. 'Laugh' moves can be seen in all meetings, but project and work had significantly fewer (3% and 2%) than discussion (6%), role-playing (7%) and chatting (11%).

The discussion meeting does not show any significant occurrence of 'check' or 'proposal', but it is the only meeting where the 'correct' move appeared (2.3% of the time). The 'contra' move was only significantly used in discussion (2.2%) and chatting (1.4%). The work meeting is the only meeting with an significant frequency of 'instruct' moves (1.7%). The move type 'polite' only appeared at a visible frequency in the project meeting (1.1%). All together, moves, which were not present in all meetings, were also not very frequent in the meetings where they did appear. These moves also include most of the moves that were added to the DSCS scheme.

5. DISCUSSION

The simple discourse function labels of the DSCS moves annotation scheme are sufficient to display the discourse dynamics of multiparty conversation. The added moves 'contra', 'correct', 'polite' and 'out' were not frequent enough to justify such an extension of the moves schema. The moves that were given the test labels could have been labeled, without significant loss of information, using moves already in the DSCS; alternatively they could be grouped under an 'other' label. More exploration is needed to determine if, for example, the 'statement' label should be broken into several more refined labels, which more accurately describe the function of the individual statement moves. For example, statements at the ends of games might be called 'summarization' or 'conclusion' moves. However, doing so would be risking confusion between a purely functional description of a speech turn and a subjective semantic description. This would introduce a different concept into what should be a purely functional moves schema.

For an initial investigation into how much information about discourse strategies in meetings can be coded with a discourse move schema, this study already shows some promising results:

The number of speakers participating in a game as well as how often individual speakers participate in games can

accurately suggest the speaker's role in the meeting structure. This information also provides a general hierarchy of speakers—it is easy to tell how many speakers were dominant, or if the speakers had balanced participation. More sample meetings from the five meeting types need to be annotated to confirm that there is a connection between speaker involvement data and the meeting type.

As possible future work, adding a third level of structure to the move/game system, such as meeting sections by topic, could provide additional information about when and how each speaker is most involved in the discourse. A further analysis of the move types used by each speaker could also give further indices about the function of a speaker in a meeting.

One of our initial questions addressed the distribution of move types across different types of meetings. There are differences in the frequency of certain move types with respect to meeting type, but these differences are not as clear as perhaps expected. The moves scheme does not seem to reflect the more subtle differences among the meeting types. However, the introduction of a 'contra' move, which indeed was solely present in discussion, points into this direction. If the data can be labeled according to topic sections, though, it may be possible to determine whether certain meeting types contain distinct 'modules;' that is, whether each meeting type has reoccurring patterns of games that vary with respect to the section's function (brain storming, decision making, status report, etc.).

Another idea that came up was to look at the alignment of speaker interference and games. Speaker interferences occur when the turn of one speaker is interfered with turns of other speakers, speaking at the same time. At first glance, it seems that games and the phases where speakers were interrupting each other or talked at the same time are in a way aligned.

In conclusion, the DSCS move scheme is an appropriate tool to express structures of communicative group interactions in multiparty conversations. There are a large number of questions that should be asked of the labeled data. Properly examined, annotation results will add to the knowledge of meeting discourse behavior and will provide support for research in the automatic summarization and description of meetings.

11. REFERENCES

- [1] Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., "HCRC Dialogue Coding Manual," *HCRC Technical Report*, HCRC/TR-82, Edinburgh, 1996.
- [2] Burger, S., Maclaren, V., Yu, H., "The ISL Meeting Corpus: The Impact of Meeting Type on Speech Style," *Proceedings of ICSLP 2002*, Denver, 2002.
- [3] Carletta, J., Isard, A., Isard, S., Kowtko, J., Newlands, A., Doherty-Sneddon, G., Anderson, A. "The Reliability of a Dialogue Structure Coding Scheme," *Computational Linguistics*, Volume 23, Pages 13-31 , Edinburgh, 1997.
- [4] Burger, S., Costantini, E. Pianesi, F., "Communicative Strategies and Patterns of Multimodal Integration in a Speech-to-Speech Translation System", *Proceedings of Machine Translation Summit 2003*, New Orleans, 2003.